Journal of Economic Growth, 3: 53–80 (March 1998)


# Technological Change, Market Rivalry, and the Evolution of the Capitalist Engine of Growth

PIETRO F. PERETTO

*Department of Economics, Duke University, Durham, NC 27708*

In the early stages of Western industrialization, innovation was the domain of individuals who devoted their entrepreneurial talents to the development of a new product or process, typically setting up a new firm in order to take the innovation to the market. Today, commercial R&D is almost exclusively carried out by corporate laboratories affiliated with manufacturing firms. The corporate R&D lab, however, did not exist in its modern form until the late nineteenth century. The history of Western industrialization, thus, suggests that a fundamental change in the structure of incentives, and consequently in the nature and the organization of the R&D process, occurred around the turn of the century. Three questions arise. What is the nature of this change? What economic forces caused it? What are its implications? To answer these questions, I construct a model where this change is endogenous to the evolution of the economy toward industrial maturity. The change in the locus of innovation—from R&D undertaken by intventor-entrepreneurs, to R&D undertaken within established firms in close proximity to the production line—results from the interaction of market structure and technological change. This interaction captures the essence of the evolution of the capitalist engine of growth and provides an economic explanation of a "stylized fact" that has received no attention in the theoretical literature. The endogenous market structure generates dynamic feedbacks that shape the growth path of the economy and determine the structural change it undergoes, including the endogenous formation of corporate R&D labs. The evolution of market rivalry explains when and how established firms become the major locus of R&D activity.

***Keywords:*** industrialization, R&D, technological change, long-run growth, entry, market structure

**JEL classification:** E10, L16, O31, O40

## 1. Introduction

In this article, I study a fact that has not received the attention it deserves in the literature on endogenous technological change. Namely, the nature of industrial capitalism has changed quite radically in the course of time. The best illustration of this "stylized fact" is Schumpeter's discussion (1928) of the shift of focus in his own work, from the inventor-entrepreneurs of his model of competitive capitalism to the professional R&D managers that characterize his model of trustified capitalism.

In the period that economic historians label the first Industrial Revolution, innovation was the domain of individuals who, spurred by the expectation of substantial economic returns, devoted their entrepreneurial talents to the development of new goods and processes. Often, taking the new product and the new process to the market required these entrepreneurs to start up a new firm (Mowery and Rosenberg, 1989, chs. 2–3). Many historical accounts of the development of major innovations, like the Watt steam engine (Scherer, 1984, ch. 2) or the Bessemer steel process (Temin, 1964), emphasize the effort of the individual inventors who devoted their time and resources to developing the innovation.[1] In time, this picture

changed. Maddison (1982, pp. 56–57) reports that 82 percent of patents granted in the United States in 1901 were granted to individuals. In 1970, in contrast, only 21 percent of patents were granted to individuals. Today, commercial R&D is almost exclusively carried out by corporate laboratories affiliated with manufacturing firms.[2] Mowery and Rosenberg (1989, ch. 6) report that in the United States in 1985, 73 percent of total industrial R&D was performed by the private, corporate sector (although only 50 percent was financed by it, the rest being financed primarily by the federal government). In a survey study conducted between 1982 and 1984, managers of R&D labs were asked to rank the contribution of various sources of technological advance in their industries. They ranked “firms in the industry” as the most important and “individual independent inventors” as the least important (Levin, Klevorick, Nelson, and Winter, 1987; Klevorick, Levin, Nelson, and Winter, 1995). Similar results emerge from work on USPO and EPO data showing that in a majority of technological classes most patents are granted to established firms (Malerba and Orsenigo, 1995, 1996; Malerba, Orsenigo, and Peretto, 1997).[3]

The corporate R&D lab, however, did not exist in the modern form until the late nineteenth century. Scherer and Ross (1990, ch. 17), for example, report that in the 1770s and 1780s the firm of Boulton and Watt had the equivalent of an R&D lab for work on steam engines, although the scale and type of activity was not exactly what one sees today (see Scherer, 1984, ch. 2, for details). They trace the genesis of the modern corporate R&D lab in the United States to 1876, when Thomas Edison opened his R&D lab in Menlo Park and Alexander Graham Bell established a similar facility in Boston. Mowery and Rosenberg (1989, pp. 38–39) have an example that, for the purposes of this article, is rather illuminating. They report that the first Bessemer steel was made in the United States in Wyandotte, Michigan, in 1864 and that in anticipation of the problems associated with chemical variations in inputs, a chemical lab was established in Wyandotte in 1863. This was the first chemical lab established in the metallurgical sector in the United States and one of the first attached to an industrial firm. Steel users made similar decisions. Railroad companies, like the Burlington Railroad in 1876 and the Pennsylvania Railroad in 1874, established their own central testing labs to make sure that the steel met appropriate specifications. Similar stories can be told about many other manufacturing sectors in the United States and in European countries (see Mowery and Rosenberg, 1989, chs. 3–4, and the references cited therein).

According to Baumol (1993, ch. 6), the most important characteristic of corporate R&D is that it is systematic, incremental, and cumulative.[4] In addition, it implies intertemporal decision-making of the type that is typically used in economic theory to characterize the accumulation of physical capital but that does not apply to R&D decisions in the standard models of endogenous technological change (see, in particular, pp. 117–119). This different type of decision-making has important implications. In one of his most cited passages, Schumpeter (1942, p. 132) remarks that “[the entrepreneur’s] social function is already losing importance and is bound to lose it at an accelerating rate . . . innovation itself is being reduced to routine. Technological progress is increasingly becoming the business of teams of trained specialists who turn out what is required and make it work in predictable ways. The romance of earlier commercial adventure is rapidly wearing away.” Commenting on this passage, Mowery and Rosenberg (1989, pp. 61–62) claim: “Although a number of major

manufacturing firms had emerged during the late nineteenth century from the innovations of such individual inventor-entrepreneurs as Eastman, Edison, Bell, and Westinghouse, in the twentieth century innovation became too important to be left to the whims of the market and the wiles of the individual inventor. . . . The growth of research within the U.S. industry was primarily an increase in research within the firm. Independent research organizations not affiliated with manufacturing firms declined in importance during the first decades of this century."[5] Similarly, Baumol (1993, p. 115) writes: "Observation appears to confirm that in reality the innovation process has veered toward becoming yet another humdrum activity of the firm, with corporate R&D taking over a substantial portion of the field and transforming it into a preprogrammed activity."

History, therefore, suggests that a fundamental change in the structure of R&D incentives, and consequently in the nature and organization of the R&D process, occurred at the turn of the century. Three questions arise. What is the nature of this change? What economic forces caused it? What are its implications?

Recently, Thompson and Waldo (1994), Smulders and van de Klundert (1995), van de Klundert and Smulders (1997), and Peretto (1994, 1995, 1996a, 1996b) have developed a number of models of endogenous innovation that focus on corporate R&D and study its macroeconomic implications.[6] These contributions can be related to the previous literature in order to understand the historical change in the locus of innovation—from R&D undertaken by outside inventors-entrepreneurs, to R&D undertaken within established firms in close proximity to the production line. In this article, I construct a model where this change is endogenous to the evolution of the economy toward industrial maturity.

Imagine a one-sector economy where oligopolistic firms sell differentiated consumption goods to households. When profitable, firms establish in-house R&D facilities to produce a continuous flow of innovations. In symmetric equilibrium, the number of firms summarizes two dimensions of the notion of market structure—concentration and relative firm size (relative to the size of the market). This property allows use of only one variable to characterize market structure. The dynamics of the number of firms are determined by the free-entry condition that the present value of profits, net of R&D costs, equal sunk entry costs. In this environment, market structure and technological change are interdependent. At a moment in time, market structure determines market rivalry and, therefore, the returns to innovation and the R&D behavior of profit-seeking firms. In addition, the number of firms changes in response to demand and technology conditions. Market structure, therefore, is endogenous and generates dynamic feedbacks that shape the growth path of the economy and determine the structural change it undergoes, including the formation of corporate R&D labs. This setup, in my view, captures the essence of the evolution of the capitalist system and provides an interpretation of the "stylized fact" that emerges from the history of Western industrialization. Namely, the evolution of the structure of the market for manufacturing goods, in particular the endogenous change in market rivalry, drives the change in the incentives faced by economic agents and thereby explains when and how established firms become the major locus of R&D activity.

The article is organized as follows. In Section 2, I present technology and preferences. In Section 3, I define and construct the Nash equilibrium (NE) with free entry for the manufacturing sector of the economy. In Section 4, I analyze the GE dynamics of the

model and discuss the transition from one form of R&D to the other. In particular, I show that the economy converges to a stable industrial structure where entrepreneurial R&D and the formation of new firms peter out, while growth is driven by corporate R&D undertaken by established oligopolists. I then discuss the properties of this steady state and provide some comparative statics results. I conclude the section with a critical review of the setup of the model and a comparison to alternative specifications that yield similar results. This comparison highlights the fundamental insight provided by the article and suggests directions for future research.

## 2. The Model

I consider a closed economy with a fixed population $L$ of identical households who supply labor services and consumption loans in competitive labor and capital markets. Each household is endowed with one unit of labor that it supplies inelastically.

### *2.1. Households*

The typical household maximizes lifetime utility

$$U(t) = \int_t^{\infty} e^{-\rho(\tau - t)} \log C(\tau)\, d\tau, \tag{1}$$

subject to the intertemporal budget constraint that the present discounted value of expenditure, cannot be greater than the present discounted value of labor and dividend income plus initial wealth,

$$\int_t^{\infty} R(\tau)E(\tau)\, d\tau \leq \int_t^{\infty} R(\tau)[W(\tau) + D(\tau)]\, d\tau + B(t),$$

where $\rho > 0$ is the individual discount rate, $R(\tau) \equiv e^{-\int_t^{\tau} r(s)\, ds}$ is the cumulative discount factor, $E$ is per capita expenditure, $B$ is asset holding, $W$ is the wage rate, and $D$ is dividends (profits are distributed in equal shares to households). Households have symmetric preferences over a range of differentiated goods

$$C = \left[\sum_{i=1}^{N} C_i^{(\varepsilon-1)/\varepsilon}\right]^{\varepsilon/(\varepsilon-1)}, \tag{2}$$

where $\varepsilon > 1$ is the elasticity of product substitution, $C_i$ is the household's purchase of each differentiated good, and $N > 1$ is the number goods (the number of firms).

The solution for the optimal expenditure plan is well known. Households set

$$\dot{E}/E = r - \rho \tag{3}$$

and, taking as given this time-path of expenditure, maximize (2) subject to $E = \sum_{i=1}^{N} P_i C_i$.

This yields the demand schedule faced by firm $i$

$$X_i = LE \cdot \left[ P_i^{-\varepsilon} \Big/ \sum_{j=1}^{N} P_j^{1-\varepsilon} \right], \tag{4}$$

where $X_i = L \cdot C_i$ is output of firm $i$ since there are $L$ identical households. It is useful for future reference to define the market share of firm $i$. Multiplying (4) by firm $i$'s price yields

$$S_i \equiv P_i^{1-\varepsilon} \Big/ \sum_{j=1}^{N} P_j^{1-\varepsilon},$$

which is the value of firm $i$'s sales divided by the value of the industry's sales.

### *2.2. Firms*

The typical firm produces one differentiated consumption good with the technology

$$L_{Xi} = h(Z_i) \cdot X_i, \tag{5}$$

where $X_i$ is output and $L_{Xi}$ is labor employment. Unit production costs $h(Z_i) \equiv Z_i^{-\theta}$, where $0 < \theta < 1$, are an isoelastic function of the firm's knowledge stock $Z_i$. When profitable, firms establish in-house R&D facilities to produce a continuous flow of innovations. Corporate R&D is described by

$$\dot{Z}_i = \alpha K \cdot \left[ L_{Zi} + \gamma \sum_{j \neq i}^{N} L_{Zj} \right], \tag{6}$$

where $\dot{Z}_i$ is the flow of innovations generated by an R&D project employing $L_{Zi}$ units of labor for an interval of time $dt$ and $\alpha K$ is the productivity of labor in R&D as determined by the exogenous parameter $\alpha > 0$ and by the stock of public knowledge $K$. The latter captures firms' intertemporal interaction in R&D. Equation (6), in addition, allows for contemporaneous interaction in R&D. The idea is that firms affect each other not only via the standard channel of intertemporal spillovers, but also through contemporaneous interaction between researchers working on different projects at different firms. In a more general setup, this interaction may be modeled as endogenous. In this article, however, I assume it to be a characteristic of the R&D technology and represent it with the "interaction" parameter $0 \leq \gamma \leq 1$. For $\gamma = 0$ the firm's technology advances only as a function of the its own R&D effort. For $\gamma = 1$ it advances as a function of the collective effort of all firms in the industry.

### *2.3. Knowledge*

This economy accumulates two types of knowledge—private and public. When one firm generates a new idea to improve its own production process, it also generates general-purpose knowledge that other firms exploit in their own research efforts. Firms appropriate

the returns from R&D through a variety of means (Levin, Klevorick, Nelson, and Winter, 1987) but cannot prevent others from using the general-purpose knowledge that spills into the public domain. The productivity of R&D is determined by some combination of all these different sources of knowledge. A simple way of capturing these notions is to posit that an R&D project that produces $\dot{Z}_i$ units of private knowledge also generates $\dot{Z}_i$ units of public knowledge and to write

$$K = \sum_{i=1}^{N} S_i Z_i$$

stating that the technological frontier is given by the weighted average of the knowledge of all firms, with weights given by the market shares. Overall, the R&D technology (6) exhibits increasing returns to scale to public knowledge and labor, and constant returns to scale to public knowledge, the accumulated factor. This property makes a steady state with constant growth feasible.

The key to this setup is that innovations are nontrivial at the firm level but not at the industry level. It is often argued that, by its own nature, technology is nonrival across firms as well as within the firm (Romer, 1990). This is true if by *technology* one means disembodied knowledge of a general nature that can be codified and transmitted at low (possibly zero) cost. In reality, industrial technology is both excludable and rival, often to a substantial degree, due to patent laws, intellectual property rights, secrecy, tacitness, and the firm-specificity of most innovations (Levin, Klevorick, Nelson, and Winter, 1987; Dosi, 1988; Malerba, 1992; Klevorick, Levin, Nelson, and Winter, 1995). Moreover, R&D is most effective when institutionally linked with production and marketing operations. Independent research firms have access to much less detailed knowledge of the production process and the characteristics of the product and cannot develop innovations of the same quality or at the same cost as in-house R&D labs. There are, therefore, strong incentives for firms to integrate the R&D function in-house and increasing returns due to the nonrival nature of technology are largely internal to the firm, the institution where innovations are developed and applied. To the extent that R&D generates general-purpose public knowledge, in addition, there are increasing returns external to the firm in the R&D process that complement those internal to the firm.

### *2.4. Entrepreneurs*

An entrepreneur can create a new firm by running an R&D project that takes as input the public knowledge $K$, generated by incumbents' R&D activity, and develops a new differentiated product and its manufacturing process (Hence, entry implies *both* product and process innovation.). Since this operation requires labor, it entails a sunk, entry cost. Because of this cost, entrepreneurs *must* develop new differentiated products since entering an existing product line in Bertrand competition with the incumbent would lead to losses. The entrepreneur who incurs the entry cost becomes an entrant and joins the industry with some initial level of productivity. In this setup, entrants are net creators of knowledge: they create a new product and the knowledge necessary to run manufacturing operations. It is

reasonable to assume that the cost of creating the new firm is proportional to the initial level of productivity and that this level is proportional to the average productivity level in the industry, $Z$.

These assumptions capture two ideas. First, entrants face escalating entry costs because of the industry's ongoing technological advance. Second, the public knowledge stock $K$ raises the productivity of both incumbents' in-house R&D and entrants' efforts to create new product lines. The key is that the same stock of public knowledge provides spillovers in both the vertical and the horizontal dimensions. Entrepreneurs can enter the industry at the average level of productivity, but they do not compensate incumbents for the services derived from the stock of public knowledge. This fact ties the two dimensions of technological advance in one and only one stock of knowledge in the economy. This suggests that the cost of entry should be modeled as proportional to the ratio between the entrant's target level of productivity $Z$ and the public knowledge stock $K$. Since in this setup the two effects of knowledge accumulation by incumbents cancel out, I can write the entry cost in units of labor as a constant, $1/\beta$.

### *2.5. Discussion*

I should emphasize some aspects of my conceptual setup that, if misunderstood, might lead to a misinterpretation of the model and its relation to the literature. The type of cost reduction discussed above can be reinterpreted as quality improvement and therefore as product innovation. Hence, it is not useful to interpret entry costs as product R&D and say that firms (incumbents) do process R&D while entrepreneurs (entrants) do product R&D. Similarly, it is not useful to say that firms do vertical product innovation (a higher-quality version of an existing product) while entrepreneurs do horizontal product innovation (a totally new product). In a model with multiproduct firms, for example, horizontal product innovation could occur without creation of new firms (the extreme case being one monopolist supplying all products). As a matter of fact, the notion of corporate R&D discussed above applies to all three types of innovation—horizontal product innovation, vertical product innovation, and process innovation. The characterization of corporate innovation versus entrepreneurial innovation does not hinge on the distinction between vertical and horizontal innovation or between product and process innovation. These distinctions are misleading because they focus on the "innovation" as opposed to the "institution" that brings the innovation to the market. The focus of this article is on the interaction between the systematic innovative activity undertaken in-house by established firms, what I call *corporate R&D*, and the creation of new independent firms that bring to the market new products and processes, what I call *entrepreneural innovation*. Regardless of what firms and entrepreneurs bring to the market—a new product, a new process, both—the focus of the article is on the historical evolution of the "institutional arrangement" through which innovations are brought to the market.[7] Despite its formal similarity, therefore, my setup is conceptually very different from the existing literature and needs to be interpreted with some flexibility. This implies a potential cost for my choice of using the Dixit and Stiglitz (1977) framework of product differentiation and formalizing corporate R&D as cost-reduction. In my judgment, however, this cost is largely offset by three benefits: first, this setup is simple and analytically tractable;

second, the notation is already familiar to most readers; third, this specific application generalizes the implications, and thereby expands the scope, of the theory of technological change and industrial dynamics that I have developed in my previous work (Peretto 1994, 1995, 1996a, 1996b). I discuss these issues in more detail in Section 4.

## 3. Equilibrium of the Manufacturing Sector

In this section, I construct the symmetric, noncooperative equilibrium with free entry and exit for the manufacturing sector of the economy. Firms face identical production and R&D technologies and demand schedules and maximize the value of their shares.

### *3.1. Definition of Equilibrium*

$$V_i = \int_t^{\infty} R(\tau)\Pi_i(\tau)\,d\tau. \tag{7}$$

Using the cost function (5), instantaneous profits are $\Pi_i = [P_i - h(Z_i)] \cdot X_i - L_{Zi}$, where, taking the wage rate as the numeraire, $L_{Zi}$ is R&D expenditure. With perfect foresight, $V_i$ is the stock market value of the firm, the price of the ownership share of an equity holder.

I consider a symmetric Nash equilibrium (NE) in open-loop strategies. Let $a_i = [P_i(\tau), L_{Zi}(\tau)]$ for $\tau \geq t$ be firm $i$'s strategy, where $P_i(\tau)$ and $L_{Zi}(\tau)$ are the time-paths of price and R&D. This price and R&D strategy induces time-paths of production, sales, and innovation. At time $t$, firms commit to time-paths of price and R&D, taking as given the time-path of the number of firms.[8] Similarly, entrants take as given the incumbents' time-paths of price and R&D. Let $V_i(N, a_1, \ldots, a_N)$ be the value of the firm when there are $N$ firms playing strategies $(a_1, \ldots, a_N)$. At time $t$, $[N, a_1, \ldots, a_N]$ is an instantaneous equilibrium with free entry and free exit if for all $i$

$$V_i[N, a_1, \ldots, a_i, \ldots, a_N] \geq V_i[N, a_1, \ldots, a_i', \ldots, a_N] \geq 0 \tag{8}$$

and for all $N$

$$V_i[N+1, a_1, \ldots, a_{N+1}] \leq 1/\beta, \tag{9}$$

where $a_i'$ denotes that firm $i$ deviates from the optimal time-paths of price and R&D while the other firms do not deviate (see Dasgupta and Stiglitz, 1980). Condition (8) requires that the active firm maximizes the present value of net cash flow, taking as given the other firms' strategies, and that this maximized value be nonnegative. The latter inequality is the free-exit condition since the scrapping value of the active firm, the opportunity cost of incumbency, is zero. Equation (9) is the free-entry condition that the value of the entrants, net of sunk entry costs, be nonpositive.

### *3.2. Pricing and Research*

The intertemporal problem of the firm is to maximize value (7), subject to the production and R&D technologies (5) and (6), the demand schedule (4), $Z_i(t) > 0$ (the initial knowledge stock is given), $Z_j(\tau)$ for $\tau \geq t$ and $j \neq i$ (the firm takes as given the rivals' innovation paths), and $\dot{Z}_i(\tau) \geq 0$ for $\tau \geq t$ (innovation is irreversible). The current value Hamiltonian is

$$CVH_i = [P_i - h(Z_i)] \cdot X_i - L_{Zi} + q_i \cdot \alpha K \cdot \left[ L_{Zi} + \gamma \sum_{j \neq i}^{N} L_{Zj} \right],$$

where the costate variable $q_i$ is the value of the patent. The firm's knowledge stock $Z_i$ is the state variable, and R&D investment $L_{Zi}$ and the product's price $P_i$ are the control variables.

Irreversibility of research and the finite labor supply impose bounds on R&D investment in this problem, $0 \leq \int_0^N (L_{Zi} + L_{Xi})\, di \leq L$. The linear Hamiltonian yields

$$L_{Zi} = \begin{cases} 0 & \text{for } 1 > q_i \cdot \alpha K \\ L_Z/N & \text{for } 1 = q_i \cdot \alpha K, \\ \infty & \text{for } 1 < q_i \cdot \alpha K, \end{cases}$$

where $L_Z/N$ is a finite value to be specified below. The case $1 > q_i \cdot \alpha K$ implies that the value of the innovation is lower than its cost and the firm does not invest. The case $1 < q_i \cdot \alpha K$ implies that the value of the innovation is higher than its cost. This case violates the GE conditions and is ruled out. The first-order conditions for the interior solution are given by equality between the marginal benefit and marginal cost of the innovation, $1 = q_i \cdot \alpha K$, the transversality condition, $\lim_{t \to \infty} R(t) \cdot q_i(t) \cdot Z_i(t) = 0$, the constraint on the state variable, $\dot{Z}_i = L_{Zi} \cdot \alpha K$, and a differential equation in the costate variable,

$$r = \frac{\dot{q}_i}{q_i} - \frac{h'(Z_i) \cdot X_i}{q_i}, \tag{10}$$

that defines the rate of return to R&D as the ratio between revenues from the innovation and its (shadow) price plus (minus) the appreciation (depreciation) in the value of the knowledge stock. Revenues from the innovation are given by the cost reduction it yields times the scale of production to which it applies. The optimal Bertrand-Nash price strategy is

$$\frac{P_i}{h(Z_i)} = \frac{\eta_i}{\eta_i - 1}, \tag{11}$$

where $\eta_i \equiv \varepsilon - (\varepsilon - 1) \cdot S_i$ is the price elasticity of demand that firm $i$ faces and the market share $S_i$ has been defined in Section 2.[9] A Bertrand-Nash equilibrium exists only if the price elasticity of demand is larger than one. In symmetric equilibrium, this requires $N > 1$.

Consider now the interior solution where the condition $1 = q_i \cdot \alpha K$ holds, and let $g_K \equiv \dot{K}/K$ be the rate of growth of public knowledge. One can reduce the first-order conditions to

$$r + g_K = \frac{LE\alpha\theta(\eta_i - 1)S_i}{\eta_i} \cdot \frac{K}{Z_i}. \tag{12}$$

This is a perfect-foresight, no-arbitrage condition that defines the rate of to R&D for firm $i$. Specifically, revenues from the innovation must equal the cost of R&D financed by borrowing at the market interest rate (direct cost), and this must equal the return on a riskless loan, at the market rate, of the resources spent on R&D (opportunity cost). The rate of return to R&D is the market interest rate $r$ plus the term $g_K$ because the cost of R&D falls with the accumulation of public knowledge.

The NE of this symmetric knowledge accumulation game is given by the first-order conditions for all firms. The following proposition establishes a property that turns out to be very useful.

**Proposition 1:** *Assume $\theta(\varepsilon - 1) < 1$. The NE has a symmetric steady state.*

*Proof.* See the appendix.

This result allows study of the dynamics of the model by focusing on the symmetric steady state of the NE.[10] This requires some additional assumptions, discussed in the appendix, but it simplifies the analysis because it reduces the space-state of the model to only two variables (see below).

In symmetric equilibrium, (12) yields the rate of innovation

$$g_Z = g_K = \alpha \cdot [1 + \gamma(N-1)] \cdot \frac{L_Z}{N}, \tag{13}$$

where $L_Z$ is aggregate employment in R&D ("aggregate R&D") and $L_Z/N$ is average employment in R&D ("average R&D"). Using (12), the rate of return to R&D becomes

$$r_{R\&D} = \alpha \cdot \left[ \frac{LE\theta(\eta - 1)}{N\eta} - \frac{L_Z}{N} \cdot [1 + \gamma(N-1)] \right], \tag{14}$$

where $\eta = \varepsilon - (\varepsilon - 1)/N$. Below, I refer to this equation as the R&D locus.

The terms $LE/N\eta$ and $\theta(\eta - 1)$ are, respectively, the *gross-profit* and the *business-stealing* effects. These arise from firms' market interaction. Intuitively, the returns to R&D are given by the gross profits earned on a given market share and by the increase in market share achieved by the R&D project. The gross-profit effect has two components—a *scale* effect, captured by the firm's sales $LE/N$; and a *markup* effect, captured by the profit margin $1/\eta$. The gross-profit effect is decreasing in the number of firms because both the market share and the oligopoly markup are lower the larger is $N$. In contrast, the business-stealing effect is increasing in the number of firms because the potential gain of market share is proportional to the rivals' total market share. To see this, note that $\theta(\eta-1) = \theta(\varepsilon-1)(N-1)/N$, where $(N-1)/N$ is the rivals' market share. Overall, the rate of return to R&D is hump-shaped in the number of firms because the business-stealing effect, which depends on the rivals' market share, and the gross-profit effect, which depends on the firm's market share, work in opposite directions.

The term $-L_Z[1 + \gamma(N-1)]/N$ is the *intertemporal-spillover* effect. Firms do not include in the value of innovations their contribution to reducing future R&D costs. This includes the contemporaneous interaction in R&D. Intuitively, the larger the contempora-

neous interaction, the larger the intertemporal spillovers since at any moment in time more knowledge is created for a given level of R&D effort.

### *3.3. Entry*

The discussion in Section 2 allows me to posit the aggregate entry technology $\dot{N} = \beta \cdot L_N$, where $\beta > 0$ is the productivity of labor in entry and $L_N$ is the amount of labor devoted to starting up $\dot{N}$ new firms for an interval of time $dt$. Note the change of notation. I defined preferences (2) over a range of differentiated goods where the number of goods $N$ is discrete. It is, however, convenient in the following analysis to treat $N$ as a continuous variable. I could have defined preferences over a continuum of goods and used integrals, thus eliminating the need to change notation. However, the assumptions discussed in Section 2 and the definition of equilibrium given in Section 3 are much more transparent if one treats the number of firms as discrete. Thus, I chose to discuss the setup of the model with a discrete $N$ and switch now to treating it as a continuous variable.

Entry costs $1/\beta$ and produces value $V_i$. The case $V_i > 1/\beta$ yields an unbounded demand for labor in entry and is ruled out. A free-entry equilibrium requires $V_i \leq 1/\beta$, with equality whenever $\dot{N} > 0$ and strict inequality whenever $\dot{N} = 0$. This implies that in an equilibrium with entry the exit condition is not binding. This is intuitive: entry costs lead to different thresholds for entry and exit. Here the threshold value for exit, zero, is never binding since firms have the option of setting R&D to zero and the value of the firms is strictly positive.

Taking logs and time-derivatives of (7) and rearranging terms yields

$$r = \frac{\Pi_i}{V_i} + \frac{\dot{V}_i}{V_i},$$

which is a perfect-foresight, no-arbitrage condition for the equilibrium of the capital market. It requires that the rate of return to firm ownership equal the rate of return to a riskless loan of size $V_i$. The rate of return to firm ownership is the ratio between profits and the firm's stock market value plus the capital gain (loss) from the stock appreciation (depreciation).

In an equilibrium with entry, $V_i = 1/\beta$ implies $\dot{V}_i = 0$, and one obtains the instantaneous free-entry condition $r = \beta \Pi_i$. Imposing symmetry yields the rate of return to entry,

$$r_{\text{entry}} = \beta \cdot \left[ \frac{LE}{\eta N} - \frac{L_Z}{N} \right]. \tag{15}$$

The condition $V_i = 1/\beta$ implies that in an equilibrium with entry, the rate of return to entry is equal to the rate of return to firm ownership. The former is equal to the ratio between net revenues and entry costs. Alternatively, if profits are distributed to shareholders, the rate of return to firm ownership is equal to the ratio between dividends and the value of the firm. Below, I refer to this equation as the entry locus.

The terms $LE/\eta N$ and $L_Z/N$ are, respectively, the gross-profit and the *R&D-cost* effects. In equilibrium, gross profits equal entry costs. These have two components—R&D expenditures and interest payments on exogenous entry costs. The former are the costs that the entrant pays to keep the pace of progress of the industry. The latter are the costs that he pays to write off the initial start-up investment.

### *3.4. The Free-Entry Equilibrium*

I have discussed two types of investments—in-house R&D and start-up of new firms. No-arbitrage in the capital market requires that they yield equal rates of return.

**Proposition 2:** *Assume $\beta < \alpha\theta(\varepsilon - 1) < \alpha$. The following free-entry equilibrium exist:*

$$L_Z = \begin{cases} 0 & \text{for } \ 1 < N \leq N_0 \\ \dfrac{LE}{\eta} \cdot \dfrac{\alpha\theta(\eta-1)-\beta}{\alpha[1+\gamma(N-1)]-\beta} & \text{for } \ N > N_0 \end{cases}, \tag{16}$$

*where* $N_0 \equiv \alpha\theta(\varepsilon - 1) - \beta]$.

*Proof.* See the appendix

In order to keep the exposition from becoming taxonomic, and to address the issues that motivate this article, I focus on this equilibrium. (A full characterization of the set of possible equilibria is available on request.) Equation (16) expresses aggregate R&D $L_Z$ as a percentage of aggregate gross profits $LE/\eta$. There is a critical number of firms $N_0$ that identifies a no-R&D region. Let $g \equiv \theta \cdot g_Z$ be the average rate of cost reduction. This is the rate of growth of labor productivity ("growth"). Equations (13) and (16) yield

$$g = \begin{cases} 0 & \text{for } \ 1 < N \leq N_0 \\ \alpha\theta \cdot \dfrac{1+\gamma(N-1)}{N} \cdot \dfrac{LE}{\eta} \cdot \dfrac{\alpha\theta(\eta-1)-\beta}{\alpha[1+\gamma(N-1)]-\beta} & \text{for } \ N > N_0 \end{cases}. \tag{17}$$

The equilibrium rate of return to investment (R&D and entry) is

$$r = \begin{cases} \beta \cdot \dfrac{LE}{\eta N} & \text{for } \ 1 < N \leq N_0 \\ \beta \cdot \dfrac{LE}{\eta N} \cdot \dfrac{\alpha[1+\gamma(N-1)-\theta(\eta-1)]}{\alpha[1+\gamma(N-1)]-\beta} & \text{for } \ N > N_0 \end{cases}. \tag{18}$$

It is useful to emphasize the relation between the growth rate and the state variables of the model.

**Result 1:** *In free-entry equilibrium, growth $g$ is hump-shaped in the number of firms $N$ and increasing in aggregate consumption expenditure LE.*

These properties are explained by the *rivalry*, *dispersion*, and *scale* effects. These effects play a central role in this model, and it is worth discussing them in detail.

The rivalry effect—the effect of an increase in the number of firms on aggregate R&D—is captured by the slope of (16). Consider Figure 1 and recall that firms take as given the time-paths of consumption and the number of firms. At a moment in time, therefore, the only variable that adjusts to ensure equality of the rates of return to R&D and entry is aggregate

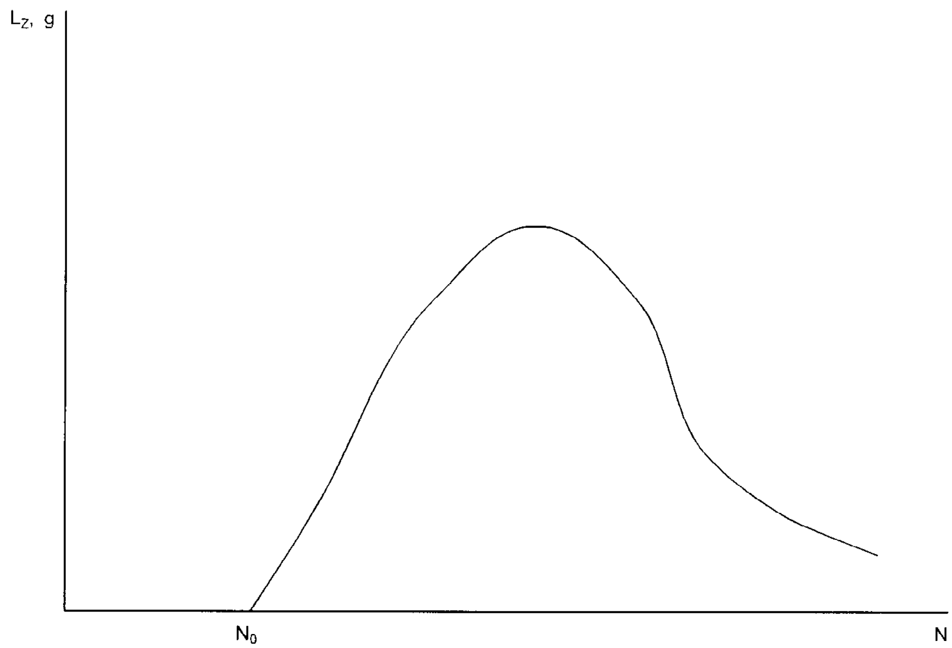


*Figure 1.*

R&D. Consider an increase in the number of firms. The rate of return to entry falls. The rate of return to R&D may fall or rise since it is hump-shaped in the number of firms. In Figure 1, the rivalry effect is initially positive and then negative. The intuition is simple. Consider Figure 2, which plots the two rate-of-return schedules in $(N, r)$ space, and recall that the condition $\alpha > \beta$ implies that labor is more productive in R&D than in entry. Let equilibrium be at the crossing of the two loci on the upward sloping portion of the R&D locus. Suppose an increase in the number of firms. The rate of return to entry is now lower than the rate of return to R&D. An increase in aggregate R&D lowers the rate of return to R&D faster than the rate of return to entry, thereby closing the gap between them and restoring equilibrium. Suppose, in contrast, that equilibrium is initially on the downward sloping portion of the R&D locus. A reduction in aggregate R&D raises the rate of return to R&D faster than the rate of return to entry and restores equilibrium.

The dispersion effect—the effect of an increase in the number of firms on growth for a given level of aggregate R&D—captures the notion that firm size and concentration of R&D resources determine the rate of productivity growth of this economy. In this model, the rate of cost reduction depends on the scale of the R&D program of the individual firm. Growth, accordingly, depends on average R&D, not aggregate R&D. The dispersion effect can be compensated by contemporaneous interaction in R&D. One can see in (17) that for $\gamma = 1$ there is no dispersion affect at all.

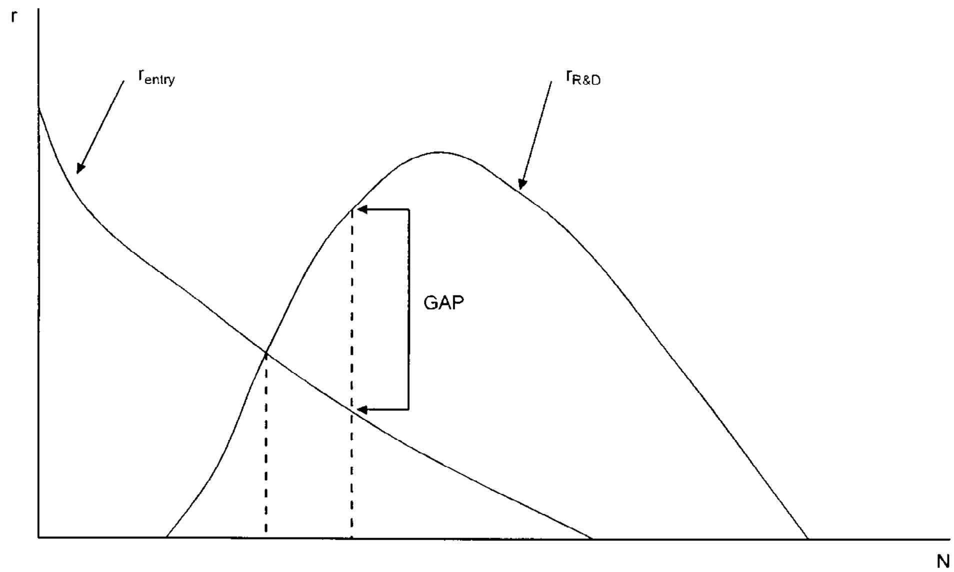


*Figure 2.*

Finally, consider an increase in aggregate demand $LE$. Both rates of return increase, but the rate of return to R&D increases more than the rate of return to entry. Hence, restoring equilibrium requires an increase in aggregate R&D. This is the standard scale effect of R&D models with increasing returns. As the size of the market increases, the rents available to innovators increase and aggregate R&D increases.

## 4. General Equilibrium, Dynamics, and the Evolution of the Capitalist Engine of Growth

In this section I characterize the transitional dynamics of the model and discuss the main results of the article. The fundamental insight is that he endogenous evolution of market structure, in particular the increase in market rivalry, drives the change in incentives that leads from a regime where growth comes from entry of new firms that bring to the market new goods and processes, to a regime where growth comes from in-house R&D that established firms undertake in order to reduce production costs and increase product quality. I conclude the section with a critical review of the setup of the model and a comparison of the formulation adopted in this article to alternative specifications that yield similar results.

### *4.1. General Equilibrium*

To determine the aggregate dynamics of the economy, I impose two GE conditions—labor market clearing and equality between the rate of return to investment and the rate of return to saving. Substituting the price strategy (11) into the cost function (5), using (4), and summing across firms yields aggregate employment in production ("aggregate production")

$$L_X = \frac{LE(\eta - 1)}{\eta}. \tag{19}$$

Labor market clearing requires $L = L_X + L_Z + L_N$, where $L_N = \dot{N}/\beta$ is aggregate employment in entry. In an equilibrium with entry, employment in all activities must be positive and $L > L_X + L_Z$ must hold. Substituting (16) and (19) into the labor market clearing condition yields

$$\dot{N} = \beta \cdot \begin{cases} L - \dfrac{LE(\eta-1)}{\eta} & \text{for } E < E_0(N) \text{ and } 1 < N \le N_0 \\ L - \dfrac{LE}{\eta} \cdot \left[(\eta-1) + \dfrac{\alpha\theta(\eta-1)-\beta}{\alpha[1+\gamma(N-1)]-\beta}\right] & \text{for } E < E_0(N) \text{ and } N > N_0 \\ 0 & \text{for } E \ge E_0(N) \text{ and } N > 1 \end{cases}, \tag{20}$$

where

$$E_0(N) = \begin{cases} \dfrac{\eta}{\eta - 1} & \text{for } 1 < N \le N_0 \\ \eta \cdot \left[\dfrac{\eta - 1}{\eta} + \dfrac{\alpha\theta(\eta - 1) - \beta}{\alpha[1 + \gamma(N - 1)] - \beta}\right]^{-1} & \text{for } N > N_0 \end{cases}.$$

The rate of return to investment is given by (18). The expenditure plan (3) yields

$$\frac{\dot{E}}{E} = \begin{cases} \beta \cdot \dfrac{LE}{\eta N} - \rho & 1 < N \le N_0 \\ \beta \cdot \dfrac{LE}{\eta N} \cdot \dfrac{\alpha[1 + \gamma(N - 1) - \theta(\eta - 1)]}{\alpha[1 + \gamma(N - 1)] - \beta} - \rho & N > N_0 \end{cases}. \tag{21}$$

The $GE$ of the economy is thus described by a system of two differential equations in $(N, E)$ space.

### *4.2. Transitional Dynamics*

Consider the phase diagram in Figure 3. The constraint $0 \le L_X \le L$ defines the feasible region of the phase plane. Using (19), this can be written

$$0 \le E \le \overline{E}(N) \equiv \frac{\eta}{\eta - 1}.$$

Above this locus labor demand exceeds the finite labor supply. The $E_0(N)$ locus defined in (20) splits the phase plane in two regions: above the locus, entry is not profitable; below it, entry is profitable. These are, respectively, the blockaded-entry and the free-entry region. Setting $\dot{E} = 0$ yields

$$E = \begin{cases} \dfrac{\rho\eta N}{L\beta} & \text{for } \ 1 < N \le N_0 \\ \dfrac{\rho\eta N}{L\beta} \cdot \dfrac{\alpha[1+\gamma(N-1)] - \beta}{\alpha[1+\gamma(N-1) - \theta(\eta-1)]} & \text{for } \ N > N_0 \end{cases}. \tag{22}$$

Expenditure increases above this locus and decreases below it. Crossing of the $\dot{E} = 0$ and $E_0(N)$ loci identifies the free-entry steady state $(N_{FE}, E_{FE})$. In the entry region to the right of $(N_{FE}, E_{FE})$ there is a saddle path leading to that point.

**Proposition 3:** *There is a unique perfect-foresight GE. If the initial number of firms is smaller than $N_{FE}$, the economy jumps on the saddle path and converges to $(N_{FE}, E_{FE})$. If the initial number of firms is larger than $N_{FE}$, the economy enters a steady state with no entry.*

*Proof.* See the appendix.

The free-entry steady state can occur to the left or to the right of $N_0$. In the former case I have a free-entry steady state with zero R&D; in the latter, I have a free-entry steady state with positive R&D. Inspection of Figure 3 suggests that the steady state $(N_{FE}, E_{FE})$ is in the R&D region if the $\dot{E} = 0$ locus evaluated at $N_0$ is below the $E_0(N)$ locus evaluated at the same point. Recalling that the definition of $N_0$ implies $\alpha\theta(\eta-1) = \beta$, one can show that this condition reduces to

$$L > \frac{\rho}{\alpha\theta} \cdot \frac{\alpha\theta(\varepsilon-1)}{\alpha\theta(\varepsilon-1) - \beta}.$$

Intuitively, the free-entry steady state is inside the R&D region if population is sufficiently large.

### *4.3. Steady-State Comparative Statics*

It is useful to describe the steady states of this model as follows. First, I construct the locus *FF* of the full-employment, value-maximizing labor allocations. This locus describes equilibria where entry is blockaded and the number of firms does not adjust endogenously to changes in parameters. (Thus, it describes the equilibrium of a model with an exogenous number of firms.) Consider the rate of return to R&D in (14). In steady state, this must equal the discount rate $\rho$. Using (19), the labor market clearing condition, and (13) yields

$$g = \begin{cases} \dfrac{\theta[1+\gamma(N-1)]}{\theta + 1 + \gamma(N-1)} \cdot \left[\dfrac{L\alpha\theta}{N} - \rho\right] & \text{for } \ 1 < N < \overline{N} \\ 0 & \text{for } \ N \ge \overline{N} \end{cases}. \tag{23}$$

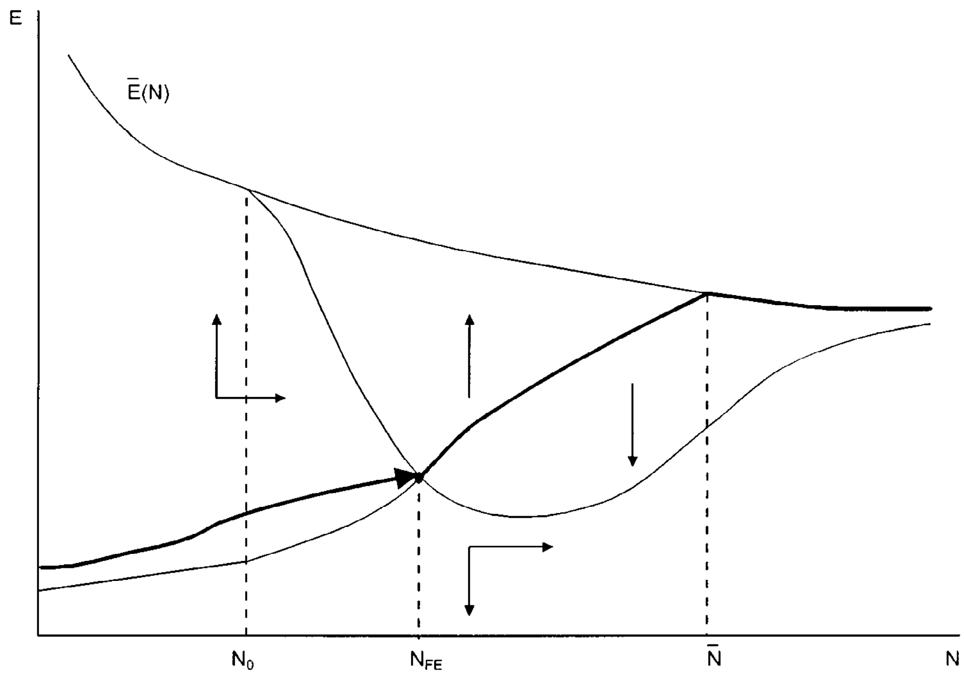


*Figure 3.*

Next, I construct the locus *NN* of the no-arbitrage, value-maximizing labor allocations. This locus describes equilibria where the number of firms adjusts endogenously to changes in parameters. The no-arbitrage condition for the rates of return to R&D and entry yields the R&D equation (16). Using (22) and (13) yields

$$g = \begin{cases} 0 & \text{for } \ 1 < N \leq N_0 \\ \dfrac{\theta\rho}{\beta} \cdot \dfrac{[1+\gamma(N-1)] \cdot [\alpha\theta(\eta-1)-\beta]}{1+\gamma(N-1)-\theta(\eta-1)} & \text{for } \ N > N_0 \end{cases} . \tag{24}$$

These equations, depicted in Figure 4, give two loci in $(N, g)$ space.

The *NN* locus is hump-shaped, initially increasing, then decreasing, but bounded from below. The *FF* locus is everywhere decreasing. The key to these loci is that for a given allocation of labor to R&D there is a tradeoff between the number of firms and growth. The *FF* locus identifies this tradeoff in the labor market. Suppose an increase in the number of firms. Equation (14) implies that the rate of return to R&D stays equal to $\rho$ if aggregate production rises. This implies a fall in aggregate R&D and, because the number of firms has risen, an even larger fall in average R&D. As a consequence, growth falls. The *NN* locus identifies the tradeoff in the capital market. When the number of firms rises, (22) and (19) require that aggregate production rise to keep the rate of return to R&D and entry equal

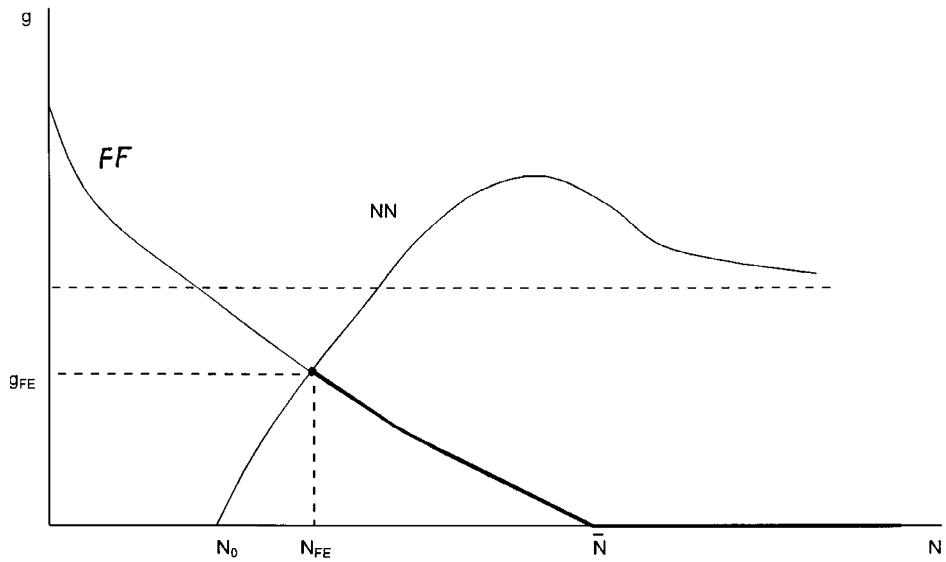


*Figure 4.*

to $\rho$. Equation (16), on the other hand, implies that aggregate R&D rises when aggregate production rises. In addition, the rivalry effect implies that the initial increase in the number of firms leads to an increase in aggregate R&D. Initially, these two effects are more than sufficient to make up for the fall in average R&D caused by the larger number of firms, and, therefore, growth rises with the number of firms. Eventually, the dispersion effect dominates, and further increases in the number of firms lead to lower growth.

A discussion of the comparative statics effects of the parameters is beyond the scope of this article. (Results are available on request.) It is, however, useful to discuss in some detail an example. Consider the effects of the opportunity parameter $\alpha$. Equation (3.10) states that the increase in the productivity of labor in R&D leads to an increase in aggregate R&D. Labor market clearing implies that aggregate production falls, and to keep all rates of return equal to $\rho$ the number of firms must fall. The intuition for the fall in the number of firms is the *escalation* effect (Sutton, 1991): better technological opportunity leads firms to spend more on R&D, but R&D expenditures are sunk costs that make entry and incumbency more costly and labor more scarce for production. These forces explain the reduction in the number of firms.[11] The effect on growth depends on the balance between the first-order effect of $\alpha$, which increases growth for any level of aggregate R&D, and the second-order effect, the endogenous adjustment in the number of firms and in the allocation of labor to R&D. Here, aggregate R&D increases and the number of firms falls. Hence, average R&D increases and growth rises.

This mechanism leads to the even more interesting results on population size $L$.

**Result 2:** *Population size has a threshold effect on growth, when the number of firms becomes sufficiently large to "trigger" R&D spending. Beyond that point, population size has a positive effect on growth. As the number of firms becomes very large, population size has a negative effect on growth that vanishes asymptotically.*

To see the global effect of population size, it is sufficient to notice that the *NN* locus does not depend on $L$ so that as population increase the *FF* locus shifts out and traces the *NN* locus. An increase in population implies that the resource base of the economy is larger. One expects both aggregate R&D and production to rise. Suppose that the extra labor force initially goes into production and that the number of firms rises. Although both these changes drive up aggregate R&D, via the rivalry and the scale effects, the dispersion effect might be so strong that growth falls. This *crowding-in* effect is the intuition for the downward-sloping portion of the NN locus: in an already crowded market a further enlargement leads to a dominant dispersion effect and a fall in growth. As the number of firms becomes very large, these effects cancel out, and growth is asymptotically independent of population size.[12]

### *4.4. Interpreting the Transition: Industrial Structure and the Nature of R&D Competition*

The phase diagram in Figure 3 allows me to address the issues raised by the historical evidence discussed in the introduction.

**Result 3:** *The economy starts out in the no-R&D region where existing firms do not undertake in-house R&D. When a critical number of firms have entered the market, established firms begin investing in R&D. The economy then converges to a steady state where no new firms are created but where productivity grows at a constant rate due to the R&D activity of established firms.*

What happens to the economy's rate of growth along this transition? At a moment in time, the behavior of the growth rate $g$ is given by (17). Along the transition, aggregate demand $LE$ and the number of firms $N$ increase. The behavior of the growth rate, therefore, depends on the interaction of the rivalry, dispersion, and scale effects. Decomposing (17), growth depends on aggregate demand $LE$ and a term that summarizes the effects of the number of firms,

$$f(N) \equiv \alpha\theta \cdot \frac{1 + \gamma(N-1)}{N} \cdot \frac{\alpha\theta(\eta - 1) - \beta}{\alpha[1 + \gamma(N-1)] - \beta}.$$

This term is hump-shaped because the rivalry effect is positive and initially dominates over the dispersion effect. Eventually, the rivalry effect turns negative and $f(N)$ goes to zero as the number of firms grows large. The two cases represented in Figure 5 are possible. In the first case, the scale effect is very strong and dominates over the rivalry and dispersion affects. The rate of growth is initially zero. It then increases throughout the transition and approaches the steady-state value from below. In the second case, the rivalry and dispersion effects, whose balance eventually turns negative, dominate over the scale effect. The rate of

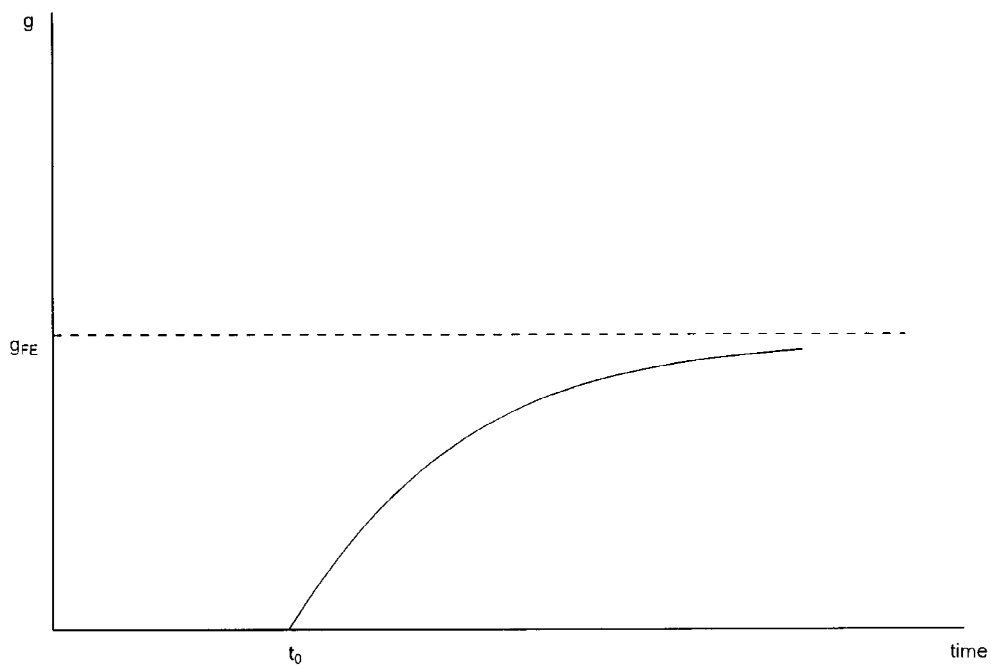


*Figure 5a.*

growth is initially zero and then increases for a while, overshooting the steady-state level. Eventually, it starts to decrease and approaches the steady-state value from above.

The key to these dynamics is the following mechanism. The R&D behavior of forward-looking firms depends on expected market rivalry because this determines the future returns to innovation. Market rivalry, on the other hand, is endogenous. In particular, the entry behavior of forward-looking firms depends on the expected R&D behavior in the post-entry equilibrium. In an environment where firms live forever and develop innovations in-house, R&D expenditure is one component of the firm's total cost and increases the cost of incumbency. It is useful to be specific here. Incumbents take as given the time-path of entry, and entrants take as given incumbents' time-paths of price and R&D. Incumbents and entrants have rational expectations and, with perfect foresight, correctly anticipate these time-paths. The symmetric free-entry equilibrium, then, can be represented with the simple feedback rule (16) that yields aggregate R&D as a function of aggregate demand and the number of firms in the market. This rule describes how, at a moment in time, current demand and market structure determine the rate of growth of the economy and its response to changes in the fundamentals and policy variables. To assess the long-run effects of these changes, one needs to track the endogenous evolution of aggregate demand and the number of firms. This point leads to the main results discussed in this section. In the long run, the number of firms, productivity growth, aggregate demand, and the interest rate

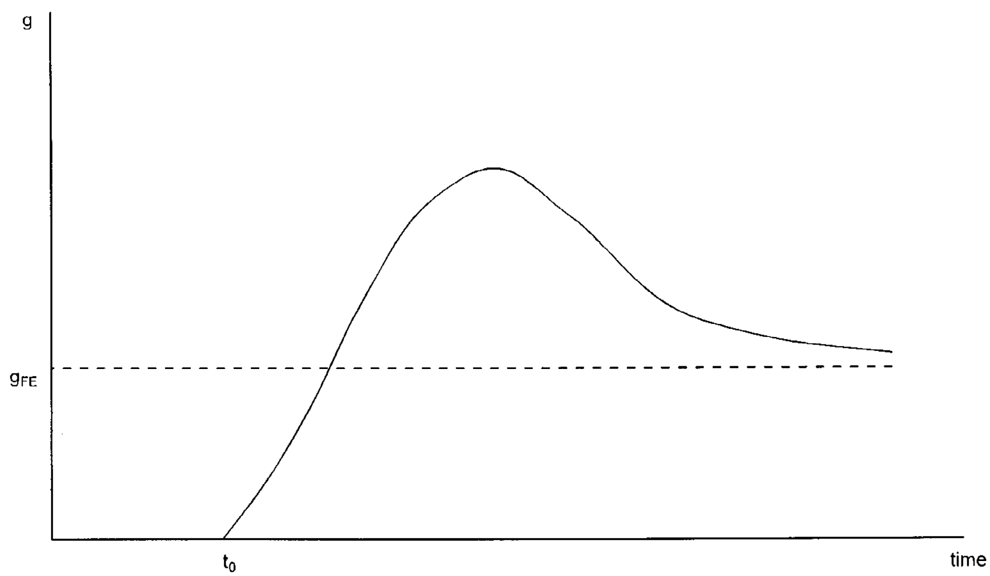


*Figure 5b.*

are endogenous and jointly determined. The number of firms, in particular, is a sufficient statistic for the state of competition and therefore for the relation between returns to R&D and returns to entry. The time-path of the number of firms, thus, predicts the timing and the shape of the transition from growth led by entrepreneurial R&D to growth led by corporate R&D.

This analytical structure allows me to tell the following story. The economy starts out with a small range of consumption goods, each one supplied by a single firm. Households like variety and buy all available consumption goods. There are, therefore, high returns from bringing new goods to the market. Entry is costly, and entrepreneurs compare the present value of profits from introducing a new good to the entry cost. Once in the market, firms live forever and engage in price competition. When profitable, they establish in-house R&D facilities. This event occurs when a sufficiently large number of firms have entered the market and rivalry has become sufficiently tough. The business-stealing effect is critical here: only when the number of firms in the market is large, it pays off for incumbents to invest in R&D in order to reduce costs, offer lower prices, and steal market share. The economy, therefore, undergoes a change in the structure of incentives. Differently from entry, in addition, I have assumed that the process of in-house innovation is self-sustaining. As firms invest in R&D, they contribute to the pool of public knowledge and reduce the cost of future R&D. These intertemporal spillovers allow the economy to grow at a constant rate in steady state. This is reached when entry peters out and the economy settles into a stable industrial structure.[13]

### *4.5. Alternative Specifications*

It is useful to compare the specification that I adopted in this article with two alternative ones that yield qualitatively similar results. The first one is straightforward. Although I tell the story in terms of consumption goods, nothing substantial changes if my oligopolistic firms do not sell differentiated consumption goods to consumers but supply specialized inputs to producers of consumption goods. Formally, one simply defines the utility index (2) as the production function for a final good that is sold by a competitive firm. This production function is thus defined over a range of nondurable intermediate inputs supplied by a number of specialized suppliers. One advantage of this formulation, one might argue, is that the expansion of the range of specialized good affects the output of the final good that grows even in the absence of in-house cost-reducing R&D in the intermediate sector. Allowing for this additional source of growth is surely interesting, above all if one wants to check the aggregate empirical implications of the model. It is not necessary, however, to make the main point of this article. Again, the key to my story is the endogenous evolution of market structure and its effects on the incentives to entry and in-house R&D. Whether entry leads only to growth of utility or also to growth of physical output is irrelevant since the main mechanism remains the same.

Similarly, nothing substantial changes if my oligopolistic firms—whether sellers of consumption goods or suppliers of intermediate goods—do not engage in cost-reducing R&D but in quality-improving R&D. Quality improvement and cost reduction are formally equivalent when one considers the quality of the good as determining the cost at which the good delivers its services (Spence, 1984; Tirole, 1988). Thus, although I focus on cost reduction, I can rewrite the model in terms of quality improvement and interpret it as a smooth version (quality being a continuous variable) of the quality-ladder model with an endogenous variety of goods (number of firms). One advantage of this formulation, one might argue, is that it makes clear that it is not the switch from product innovation to process innovation that matters in this article. Whether growth comes from innovations that improve labor productivity in the same sector where they are originated or in some downstream sector is irrelevant to my story (see Scherer, 1984, for an empirical definition of product and process R&D along these lines). What really matters is whether innovations are carried out by established producers or by outsiders that set up new firms in order to bring the innovations to the market. Established firms make decisions based on the marginal value of knowledge.[14] Entrants, in contrast, base their decisions on the value of the firm. In particular, they do so by taking into account that once they join the industry, they have to keep improving the product, or reduce the cost, in order to keep up with the competition. This is a crucial point because R&D expenditure is a sunk cost that increases the cost of incumbency. It is the interaction between incentives to entry and incentives to in-house R&D that shapes the transition path depicted in Figure 3. As argued, this interaction is captured by the endogenous evolution of market structure.

## 5. Conclusion

History suggests that new industries are borne out of radical innovations that come about in a crude form and whose arrival is driven by events outside the domain of economic forces. Crafts (1996), for example, argues that *macroinventions*, like the invention of the internal combustion engine, the light bulb, the transistor, or the personal computer, all inventions that started some of today's major "technologies" or "industries," should be described as unpredictable events that to a large extent originate from outside the economic system. Once the basic invention appears, however, economic forces take over and drive the development of the technology through the systematic implementation of *microinventions*. If one imagines that the model developed in this article describes industries that at some moment in time start out under similar conditions, then each industry goes through a *life cycle* described by the dynamics discussed in the previous section. To the extent that industries are similar, their life cycles are synchronized.

In many industrial sectors a major change in the locus of innovation occurred at the turn of the nineteenth century and in the early twentieth century. In those years, a large number of basic inventions appeared that either started entirely new industries or revolutionized existing ones. These macroinventions appeared as an *artifact* and a *technique*, a specific product and the process to make it, and a *knowledge base* that provided the framework for further development of the technology through minor but systematic improvements (see Mowery and Rosenberg, 1989, chs. 3–4, and the references cited therein for a number of examples). When the basic idea of a truly new good appears—say, the idea of the personal computer or of the integrated circuit—a revolution occurs: something fundamentally new has been invented. Technologically, much of the development of the original idea is just a refinement of the original version of the product or process. According to Rosenberg (1982, ch. 3), however, this is where the contribution of the new idea to productivity and living standards comes from. Economically, the microinventions that perfect the original macroinvention are crucial (although they might be, and often are, perceived as technologically marginal). The question, then, is where do these microinventions come from and at what rate do they arrive. This article provides an answer. The profit, often phenomenal, made by the original monopolist attracts entry. As more and more entrepreneurs enter the market, the industry becomes more competitive. As market rivalry becomes tougher, established producers find it profitable to start exploring the technology through in-house R&D. The economy, then, experiences a transition from entrepreneurial to corporate R&D.

It is straightforward to generalize this story to interpret the evidence discussed in the introduction. The cluster of macroinventions that occurred in the late nineteenth century and early twentieth century gave birth to industries whose dynamics are similar and synchronized because the basic economic mechanism driving their development is the same. This suggests the next question on the research agenda. Why did these macroinventions appear in a cluster? To the extent that macroinventions are not events that can be entirely explained or predicted by looking at strictly economic forces, answering this question requires a careful analysis of the history of each technology in order to identify the scientific and technological forces that determined its appearance in the first place. Perhaps it would help to think of macroinventions as the solutions to some critical problems that constrain determined

human activities, needs or desires (like the desire of flying and the consequent invention of the airplane). A better understanding of the technological and economic problem that each macroinvention solved, together with a better understanding of the basic motivation that spurred the inventor and of the basic constraints he had to face, would help explain the clustering of major inventions that appeared at the turn of the century and that originated many of today's industries.

## Appendix: Proofs of Propositions

### *Proof of Proposition 1*

Substituting (11) into (10) and integrating forward yields

$$q_i(t) = \int_t^\infty R(\tau)[LE\theta(\eta_i - 1)S_i/\eta_i Z_i]\, d\tau,$$

which defines the value of the innovation as the present discounted value of its contribution to current and future profits. The price strategy (11) yields a downward-sloping relationship between the product's price $P_i$ and the firm's knowledge stock $Z_i$. Holding constant the knowledge stocks of all firms other than $i$, as the knowledge of firm $i$ goes to infinity, the firm's price goes to zero and the market share $S_i$ goes to one. This yields $\eta_i = 1$, and the value of the innovation goes to zero. In contrast, as the knowledge stock of firm $i$ goes to zero, the firm's price goes to infinity and the firm's market share goes to zero. This yields $\eta_i = \varepsilon$ and the value of the innovation depends on the ratio between the market share and the knowledge stock,

$$\lim_{Z_i \to 0} \frac{S_i}{Z_i} = \lim_{Z_i \to 0} \frac{Z_i^{\theta(\varepsilon-1)-1}}{\sum_{i=1}^N Z_i^{\theta(\varepsilon-1)}} = \begin{cases} +\infty & \text{for } \theta(\varepsilon - 1) < 1 \\ 0 & \text{for } \theta(\varepsilon - 1) > 1 \end{cases} .$$

If the value of the innovation is decreasing in the firm's knowledge stock, the $NE$ is symmetric and stable in the sense that if firms start out with different knowledge stocks, they converge to symmetry. Hence, a sufficient condition for stability is that the market share as a function of the knowledge stock satisfies the Inada-type condition that the marginal gain of market share is infinite when the firm's knowledge stock is close to zero. This intuition explains why the value of the innovation goes to zero when the knowledge stock is infinite: the firm becomes a global monopolist, and, as its market share approaches one, the marginal gain in market share approaches zero. The condition $\theta(\varepsilon - 1) < 1$, therefore, is sufficient for stability because it rules out situations where the firm's market share exhibits increasing returns to the knowledge stock. Now, there are two ways to support a symmetric $NE$ at all times:

- Suppose that innovations are not firm-specific and can be traded in the patent market. Like in partial equilibrium models of investment in physical capital, this property implies that the firm's knowledge stock can change discretely at a point in time. This, however, does not mean that the aggregate knowledge stock can do so. Increases in the aggregate

knowledge stock can be achieved only by allocating labor to R&D, and this activity is always constrained by the finite labor endowment. Since firms can trade knowledge, the industry-level transitional dynamics reduce to a jump to the steady state.

- Suppose that innovations are firm specific or cannot be traded in the patent market. Since innovations have to be produced in-house, the firm's knowledge stock cannot change discretely since there is an upper bound on R&D implied by the given labor supply. If firms have different initial knowledge stocks, there is convergence in time to the symmetric steady state. In order to have symmetry at all times one needs to assume identical initial conditions. This assumption and the characterization of entry discussed in Section 2 yield that the industry is always in symmetric equilibrium.

***Proof of Proposition 2***

The free-entry equilibrium can be represented in a diagram with the rate of return $r$ on the vertical axis and aggregate R&D $L_Z$ on the horizontal axis. The restriction $\theta(\varepsilon - 1) < 1$ implies that $\theta(\eta - 1) < 1$ for all $N > 1$. This implies that the R&D locus cuts the horizontal axis at a point to the left of that where the entry locus cuts it. The parameters $\alpha$ and $\beta$ determine the slopes of the R&D and entry loci. A sufficient condition of the two loci to cross for positive values of the rate of return $r$ is $\alpha > \beta$, which implies that the productivity of labor in R&D is higher than the productivity of labor in entry. This implies that the R&D locus is steeper than the entry locus since $\alpha[1 + \gamma(N - 1)] > \beta$ for all $N > 1$. To have an equilibrium with positive R&D, the R&D locus must have a higher intercept than the entry locus. This requires $\alpha\theta(\eta - 1) > \beta$. If this condition is violated, the two loci cross for a negative value of aggregate R&D, the nonnegativity constraint on R&D is binding, and $L_Z = 0$. Hence, the condition $\alpha\theta(\varepsilon - 1) > \beta$ must hold, otherwise R&D is always zero. Therefore, an equilibrium with entry and R&D (possibly zero) exists if $\beta < \alpha\theta(\varepsilon - 1) < \alpha$.

***Proof of Proposition 3***

All points on the $\dot{E} = 0$ locus in the blockaded-entry region are steady states. These are the blockaded-entry steady states with positive R&D. Crossing of the $\dot{E} = 0$ and $\bar{E}(N)$ loci identifies the point $(\bar{N}, \bar{E})$, where $\bar{N} \equiv L\alpha\theta/\rho$. All points on the $\bar{E}(N)$ bound to the right of $(\bar{N}, \bar{E})$ are blockaded-entry steady states with zero R&D since they imply that all labor is employed in production. The stable manifold of the system is the union of the saddle path in the entry region, the portion of the $\dot{E} = 0$ locus in the blockaded-entry region below the $\bar{E}(N)$ bound, and the portion of the $\bar{E}(N)$ bound to the right of $(\bar{N}, \bar{E})$. Paths below the stable manifold eventually yield zero production. In this case, the rate of return in (18) falls to zero and the transversality condition for firms' optimal R&D plans is violated (see Section 3). Paths above the stable manifold eventually yield that demand for labor exceeds labor supply. Stopping on the labor supply constraint is not feasible since it violates the Euler's equation (21).

## Acknowledgment

I thank an anonymous referee for valuable comments.

## Notes

1. Cardwell (1995) provides a detailed account of the development of technology from ancient times to today. Over so long a period of time, it is impressive to see the change in the way technology progresses. Mowery and Rosenberg (1989) argue that what characterizes the post-1859 period, which they suggest be called the "second Industrial Revolution," is the different nature of the advance of technology—in particular, its increasing scientific content and its organization.
2. Scherer (1984, ch. 3) constructs an input-output matrix of interindustry technology flows in the U.S. economy based on 1974 data. He finds that the manufacturing sector, which accounts for 95 percent of all company-financed R&D, exports to nonmanufacturing half of the R&D it originates. Of the half that it does not export, 57 percent is process R&D—that is, R&D performed in the same industry of use. In contrast, the amount of R&D originated, financed, and used within the nonmanufacturing sector is only 7 percent of the amount imported from manufacturing. Scherer concludes: "Evidently, if one has comparative advantage at doing industrial R&D, as most larger manufacturing enterprises have, there is a tendency toward considerable self-sufficiency in developing one's own specialized production technology" (p. 51).
3. In addition, in most cases when a new innovator appears on the scene of a technological class, it is a case of lateral innovation—that is, an established firm active in a neighboring class that diversifies into the class in question (Malerba and Orsenigo, 1995, 1996).
4. It is also, most of the time, rather unglamorous, as the numerous anecdotes and the systematic evidence discussed by Mowery and Rosenberg (1989, chs. 2–4) suggest. For example, the lab established at Wyandotte had the primary task of testing the quality of inputs, not inventing new ones.
5. For example, they report that the share of employment of scientific professionals in independent R&D organizations was 15.2 percent of the total in 1921 and 6.9 percent in 1946 (Mowery and Rosenberg, 1989, p. 83, Table 4.7).
6. These papers incorporate into general equilibrium (GE) growth models the market structure and innovation approach of the industrial organization (IO) literature. See Peretto (1996a, 1996b) for a detailed discussion. Kamien and Schwartz (1982), Baldwin and Scott (1987), Cohen and Levin (1989), and Scherer and Ross (1990, ch. 17) review the IO literature, theoretical, and empirical.
7. A corollary to this discussion is that it is important to interpret $N$ as the number of firms, not the number of industries, like it is common place in the literature that uses the Dixit and Stiglitz (1977) framework to generate models of monopolistic competition.
8. One might argue that this degree of commitment is unrealistic. However, choosing open-loop strategies is necessary to solve the model in closed form and study its aggregate properties. Feedback strategies are more realistic because they are subgame perfect, but I cannot solve the model in that case.
9. Note that the price elasticity of demand is not equal to the elasticity of product substitution. This approximation, regularly used in the literature, is based on the assumption that firms are atomistic and take the industry price index as given. Here firms take into account their effect on the industry price index. Thus, there is oligopolistic rivalry. This generates the additional term that captures the fact that the demand curve faced by each firm becomes more elastic the smaller the firm's market share (see Beath and Katsoulacos, 1991, ch. 3). The industry becomes more competitive as more firms enter. Monopolistic competition emerges as the limit for a number of firms that tends to infinity.
10. This is the steady state of the partial equilibrium model only, not the steady state of the general equilibrium model. The latter is studied in Section 4. For a discussion of the properties of this class of capital accumulation games, see Spence (1984), Fershtman and Meuller (1984), Tirole (1988, ch. 8), and Fudenberg and Tirole (1991, ch. 13).
11. This is a comparative statics result only and does not take into account the dynamics of the model. Economies with different parameter $\alpha$ converge to different free-entry steady states characterized by different numbers of firms. Thus, in the case discussed in the text, the economy with the larger $\alpha$ converges to the free-entry steady state with the smaller number of firms. The same economy's response to a change in the parameter

$\alpha$ is a different question. The model, in fact, exhibits hysterisis: temporary changes in parameters can have permanent effects, since in free-entry equilibrium the exit condition is never binding (see Section 3). Thus, an increase in $\alpha$ that calls for a reduction in the steady state number of firms does not lead to firms exiting the market. Rather, the economy enters immediately a steady state with no entry and adjusts its growth rate only.

12. I should emphasize one implication of this result: the sign and magnitude of the scale effect depend on the specific stage of industrialization. That is, they depend on market structure. In Peretto (1996a, 1996c), I study in detail the scale effect and the role of population growth in models of this class.
13. The model allows for entrepreneurial innovation in the long run only in population grows. This empirical prediction fits two well-established facts in IO: larger countries have a larger number of firms; entry rates are positively correlated to the rates of expansion of markets. In Peretto (1996c), I show that if population grows at a constant rate, the steady-state rate of entry is equal to the growth rate of the labor force, while the rate of cost reduction does not depend on the size or the growth rate of the labor force. Asymptotically, this result obtains in this model as well.
14. If one interprets in-house product innovation as bringing to the market the next generation of a good that replaces the old one, as in quality-ladder models, it is immediate to see that established firms internalize the creative-destruction effect discussed by Grossman and Helpman (1991) and Aghion and Howitt (1992). In the deterministic environment of this article, this is just the Arrow replacement effect since all vertical innovations are carried out by incumbents (see Tirole, 1988, ch. 10). Thus, my setup does not rule out product obsolescence. It just internalizes within the firm the decision about the rate at which products are replaced (see Thompson and Waldo, 1994, for a good discussion of this point).

## References


Aghion, P., and P. Howitt. (1992). "A Model of Growth Through Creative Destruction." *Econometrica* 60, 323–351.

Baldwin, W., and J. Scott. (1987). *Market Structure and Technological Change*. New York: Harwood Academy.

Baumol, W. (1993). *Entrepreneurship, Management, and the Structure of Payoffs*. Cambridge: MIT Press.

Beath, J., and Y. Katsoulacos. (1991). *The Economic Theory of Product Differentiation*. Cambridge: Cambridge University Press.

Cardwell, D. (1995). *The Norton History of Technology*. New York: Norton.

Cohen, W., and R. Levin. (1989). "Empirical Studies of Market Structure and Innovation." In Richard Schmalensee and Robert Willig (eds.), *Handbook of Industrial Organization*. Amsterdam: North Holland.

Crafts, N. (1996). "The First Industrial Revolution: A Guided Tour for Growth Economists." *American Economic Review Papers and Proceedings* 86, 197–201.

Dasgupta, P., and J. Stiglitz. (1980). "Industrial Structure and the Nature of Innovative Activity." *Economic Journal* 90, 266–293.

Dixit, A., and J. Stiglitz. (1977). "Monopolistic Competition and Optimal Product Diversity." *American Economic Review* 67, 297–308.

Dosi, G. (1988). "Sources, Procedures, and Microeconomic Effects of Innovation." *Journal of Economic Literature* 36, 1120–1171.

Fershtman, C., and E. Mueller. (1984). "Capital Accumulation Games of Infinite Duration." *Journal of Economic Theory* 33, 322–339.

Fudenberg, D., and J. Tirole. (1991). *Game Theory*. Cambridge: MIT University Press.

Grossman, G., and E. Helpman. (1991). *Innovation and Growth in the Global Economy*. Cambridge: MIT University Press.

Kamien, M., and A. Schwartz. (1982). *Market Structure and Innovation*. Cambridge: Cambridge University Press.

Klevorick, A., R. Levin, R. Nelson, and S. Winter. (1995). "On the Sources and Significance of Interindustry Differences in Technological Opportunities." *Research Policy* 24, 185–205.

Levin, R., A. Klevorick, R. Nelson, and S. Winter. (1987). "Appropriating the Returns from Industrial R&D." *Brookings Papers on Economic Activity* 3, 783–820.

Maddison, A. (1982). *Phases of Capitalist Development*. Oxford: Oxford University Press.

Malerba, F. (1992). "Learning by Firms and Incremental Technical Change." *Economic Journal* 102, 845–859.

Malerba, F., and L. Orsenigo. (1995). "Schumpeterian Patterns of Innovation." *Cambridge Journal of Economics* 19, 47–65.

Malerba, F., and L. Orsenigo. (1996). "On New Innovators and Ex Innovators: An Empirical Analysis of the Patterns of Innovative Natality and Mortality." Manuscript, Bocconi University.

Malerba, F., L. Orsenigo, and P. Peretto. (1997). "Persistence of Innovative Activities, Sectoral Patterns of Innovation, and International Technological Specialization." *International Journal of Industrial Organization* 15, 801–826.

Mowery, D., and N. Rosenberg. (1989). *Technology and the Pursuit of Economic Growth*. Cambridge: Cambridge University Press.

Peretto, P. (1994). "Essays on Market Structure and Economic Growth." Ph.D. dissertation, Yale University.

Peretto, P. (1995). "Cost Reduction, Entry, and the Dynamics of Market Structure and Economic Growth." Working Paper 95/48, Department of Economics, Duke University.

Peretto, P. (1996a). "Industrialization, Technological Change, and Long-Run Growth." Working Paper 96/22, Department of Economics, Duke University.

Peretto, P. (1996b). "Sunk Costs, Market Structure and Growth." *International Economic Review* 37, 895–923.

Peretto, P. (1996c). "Technological Change and Population Growth." Working Paper 96/28, Department of Economics, Duke University.

Romer, P. (1990). "Endogenous Technological Change." *Journal of Political Economy* 98, S71–S102.

Romer, P. (1996). "Why, Indeed, in America? Theory, History, and the Origins of Modern Economic Growth." *American Economic Review Papers and Proceedings* 86, 202–206.

Rosenberg, N. (1982). *Inside the Black Box*. Cambridge: Cambridge University Press.

Rosenberg, N. (1994). *Exploring the Black Box*. Cambridge: Cambridge University Press.

Scherer, F. (1984). *Innovation and Growth: Schumpeterian Perspectives*. Cambridge: MIT University Press.

Scherer, F., and D. Ross. (1990). *Industrial Market Structure and Economic Performance*. Boston: Houghton Mifflin.

Schumpeter, J. (1928). "The Instability of Capitalism." *Economic Journal* 38, 361–386.

Schumpeter, J. (1942). *Capitalism, Socialism, and Democracy*. New York: Harper & Row.

Smulders, S., and T. van de Klundert. (1995). "Imperfect Competition, Concentration and Growth with Firm-Specific R&D." *European Economic Review* 39, 139–160.

Spence, M. (1984). "Cost Reduction, Competition, and Industry Performance." *Econometrica* 52, 101–121.

Sutton, J. (1991). *Sunk Costs and Market Structure*. Cambridge: MIT University Press.

Temin, P. (1964). *Iron and steel in Nineteenth-Century America*. Cambridge: MIT Press.

Thompson, P., and D. Waldo. (1994). "Growth and Trustified Capitalism." *Journal of Monetary Economics* 34, 445–462.

Tirole, J. (1988). *The Theory of Industrial Organization*. Cambridge: MIT Press.

van de Klundert, T., and S. Smulders. (1997). "Growth, Competition and Welfare." *Scandinavian Journal of Economics* 99, 99–118.